Multimodal data representation and information fusion algorithm

Project ID : 630

Team composition

SL. NO	Name of the student	SRN	PHOTO
1	Yogya Chukkapalli	PES1201701144	
2	Apeksha Gaonkar	PES1201701562	
3	P Jahnavi Raman	PES1201701152	

Guide : Dr. Sanjeev G

Contents

Problem statement

To develop multimodal data representation and optimal information fusion algorithm.

Solution

We developed a dependency based multimodal data representation model using R-Vine Copula

Literature review

Paper [1]: S. Zhang, L. N. Theagarajan, S. Choi and P. K. Varshney, "Fusion of Correlated Decisions Using Regular Vine Copulas," in IEEE Transactions on Signal Processing, vol. 67, no. 8, pp. 2066-2079, 15 April, 2019, doi: 10.1109/TSP.2019.2901379.

In this paper they designed an optimal fusion rule using a regular vine copula under the Neyman–Pearson framework with binary quantizers for the parallel distributed detection system. R-Vines takes complex dependence into account for the fusion of dependent decisions.

Paper [2]: S. Zhang, B. Geng, P. K. Varshney and M. Rangaswamy, "Fusion of Deep Neural Networks for Activity Recognition: A Regular Vine Copula Based Approach," 2019 22th International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, 2019, pp. 1-7

In this paper, an R-vine based feature fusion approach was used to perform the human activity recognition using multi-modal sensor observations.The features from each modality were extracted using DNN technique and the inter-modal and cross-modal dependencies between them are captured using R-vine copula model.

Paper [3]: D. Qian et al., "Drowsiness Detection by Bayesian-Copula Discriminant Classifier Based on EEG Signals During Daytime Short Nap," in IEEE Transactions on Biomedical Engineering, vol. 64, no. 4, pp. 743-754, April 2017, doi: 10.1109/TBME.2016.2574812.

In this paper, EEG and EOG sensor data is taken to classify whether the person is awake or asleep. Bayesian classifier is used for classification after D-vine fusion of multivariate data. The advantage of the developed D-vine copula Bayesian classifier over a simple Bayesian classifier is that as a connection function of the marginal distribution and joint distribution, the D-vine copula can be selected by the most suitable type to fit the linear or non-linear dependencies of features. The connection function is not subject to the restriction of a marginal probability distribution.

Paper [4]: C. Zhang, Z. Yang, X. He and L. Deng, "Multimodal Intelligence: Representation Learning, Information Fusion, and Applications," in IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 3, pp. 478-493, March 2020, doi: 10.1109/JSTSP.2020.2987728.

This paper provides a technical review of the available models and learning techniques using deep learning methods.It provides a detailed analysis of the recent works on multimodal deep learning mainly focusing on three aspects : multimodal representations,fusing multimodal signals at various levels and multimodal applications.It focuses on the combination of vision

Objective

Motivation:

- Usage of multiple sensors to acquire information about same event in order to gain new perspectives has been of interest in the recent years due to the development in both sensor technologies and IoT.
- But the high dimensionality of data poses problem in transmitting the data, storing the data and analysing it.

Solution:

- We focus on developing a state of art model for multimodal data fusion using copula theory which can not only represent the mutual information among fused features but also the characteristics specific to that feature.
- We also addressed the problem of feature selection while constructing a copula fusion model to achieve the balance of computation efforts and classification performance.

Selection Of Fusion Algorithm

Requirement :

A fusion algorithm that can model the complex dependency between data by identifying the mutual information and also retaining information specific to that data channel.

Copula:

The application of copula theory is widespread in the fields of econometrics and finance. However, its use for signal processing applications has been quite limited.

- Copula allows separation of modeling univariate marginals from modeling the dependence structure.
- It incorporates spatial dependence which can be considered non-linear and allow for any marginal distribution to be used which are non-Gaussian.
- Easier interpretation compared to Deep Neural Networks.

Copula:

According to Sklar theory, the multi-dimensional joint distribution function can be decomposed into several unary marginal distribution functions and one multi-dimensional copula function.

For the multi-dimensional data x, the joint distribution function is F(x1, x2, . . ., xn), and the marginal distribution functions are F1(x1), F2(x2), . . ., Fn(xn). There is a unique copula function C such that

$$F(x1, x2, \ldots, xn) = C(F1(x1), F2(x2), \ldots, Fn(xn)). \quad \text{--- eqn 1}$$

When the copula function is differentiable, the joint probability density function can be expressed as

$$f(x1, x2, \ldots, xn) = (f(x1) \times f(x2) \times \cdots \times f(xn)) \times c(F1(x1), F2(x2), \ldots, Fn(xn); \quad \text{--- eqn2}$$

where f(xi) is the marginal probability density function of x, c is regarding the multidimensional copula density function .

The multi-dimensional copula density function is defined as

$$c(F_1, F_2, \ldots, F_n; \lambda) = \frac{\partial^n C}{\partial F_1 \partial F_2 \cdots \partial F_n}.$$

Selection Of Copula

Requirement :

A copula model to structure higher dimensional data.

R-Vine copula:

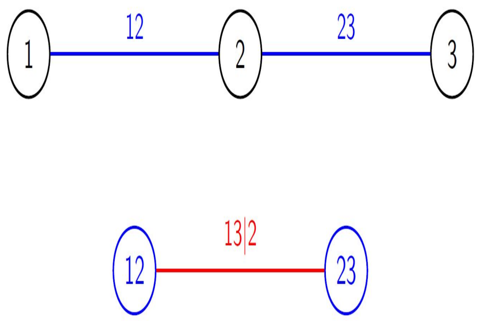
R-Vine copulas are tree-like structure that model dependencies for higher dimensional data.

There are two special categories of R-Vine based on their structure.

1. C-vine: Canonical Structure (Centric).
2. D-vine: Drawable Structure (Direct).

Rvine:

- Vine copulas are for higher-dimensional data .
- Bivariate copulas are building blocks for higher-dimensional distributions of vine copulas.
- The dependency structure is determined by the bivariate copulas and a nested set of trees.
- Vine approach is more flexible, as we can select bivariate copulas from a wide range of (parametric) families.
- The crucial part of model selection of vine copulas is to identify adequate trees and their structure.
- The number of possible regular vines on m variables is $\frac{n!}{2} * 2^{(n-2)}$
- This remains true even when the selection is restricted to the sub-classes of C- and D-vines, since there are still $n!\ /2$ different C- and D-vines in m dimensions.
- Best fit Rvine depends on :
  a. Tree structure estimation : Maximum spanning tree approach.
  b. Pair copula selection : The absolute empirical Kendall’s or AIC.
  c. Copula parameter selection: MLE or Kendall's tau.



Pair Copula construction:

The joint density can be represented as a product of pair copulas and marginal densities..
For example, let f(z1,z2,z3) be the joint density of a 3-dimensional data.

$$f(x_1,x_2,x_3) = f_{3|12}(x_3|x_1, x2) \times f_{2|1}(x_2|x_1) \times f_1(x_1)$$ --- eq1

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) \times f_2(x_2)$$ --- eq2

$$f_{3|2}(x_3|x_2) = c_{23}(F_2(x_2), F_3(x_3)) \times f_3(x_3)$$ --- eq3

$$f_{3|12}(x_3|x_1, x2) = c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \times f_{3|2}(x_3|x_2)$$ --- eq4

$$f(x_1,x_2,x_3) = f_1(x_1) \times f_2(x_2) \times f_3(x_3) \qquad (marginals)$$

$$\times c_{12}(F_1(x_1), F_2(x_2)) \times c_{23}(F_2(x_2), F_3(x_3)) \qquad (unconditional\ pairs)$$ --- eq5

$$\times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \qquad (conditional\ pairs)$$

Feature Selection

The major deciding part of the model performance is how effectively we extract information from the data.

Information extracted should have

- Taken the mutual information from data channels into consideration.
- Redundant data should be removed to avoid unnecessary processing.

Dependencies can help us extract the effective information.

- Pearson's correlation equation can be used to calculate linear dependencies.
- PP-Score metric can be used to calculate non-linear dependencies.

Pearson’s correlation equation

- The linear correlation between the features is calculated using the Pearson’s correlation equation.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

PP-Score metric

- The nonlinear dependency is calculated using PP-Score metric.
- PP-Score lies between 0 (no predictive power) and 1 (highest predictive power).
- The base line or the ground truth of this score is the score for a very naive model.
- When the target is
  - Numeric : Decision Tree Regressor is used.

    Mean Absolute Error (MAE) is calculated.

    Naive model : Predicting the median value.

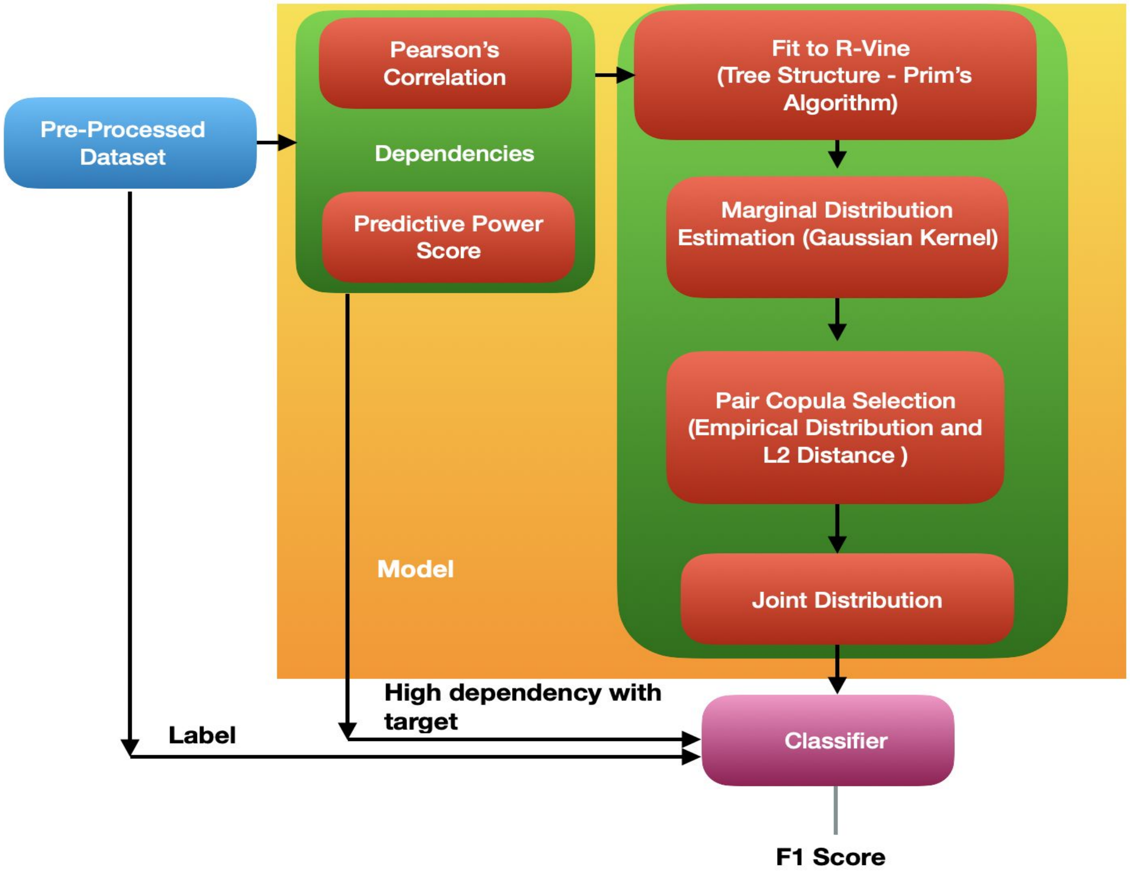
    PPS = 1 - (MAE_model / MAE_naive)
  - Categoric : Decision Tree Classifier is used.

    Weighted F1 is calculated.

    Naive model : Predicting the most common class

    PPS = (F1_model - F1_naive) / (1 - F1_naive)

Block Diagram



Dataset

STISEN:

Human Activity Recognition dataset collected from Accelerometer and Gyroscope sensors from both Phone and Watch. The sensors are 3 dimensional with x,y and z values each for a specific Activity. In total there are 6 activity classes.

TMD:

Transport Mode Detection dataset collected using 10 different sensors present in smart phone for about 5 seconds for each mode of transport. This time series data is plotted and it's mean and standard deviation values are obtained. There are 5 modes of transport.

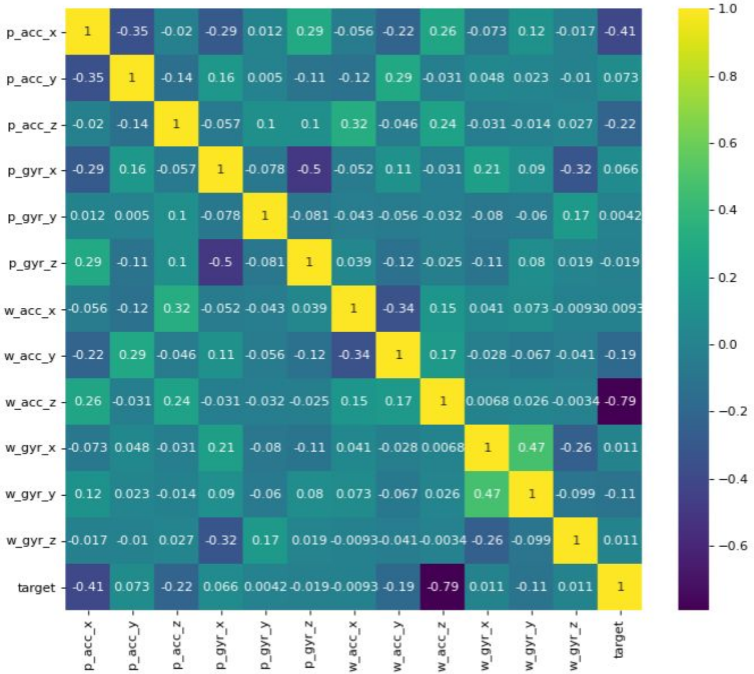
Handled null values by replacing it with the mean value of the data filtered based on class.

Dependencies heatmaps

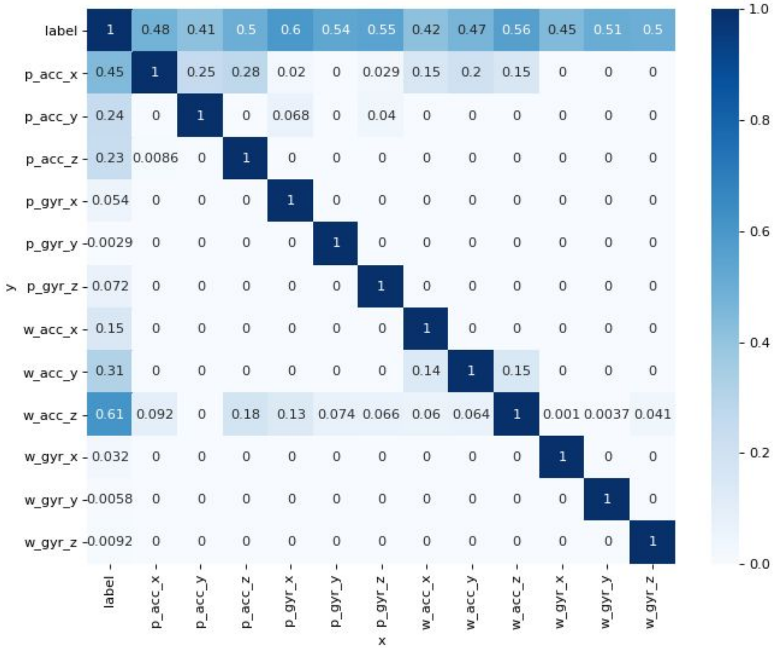
STISEN:

FIRST LEVEL FUSION

. *Correlation between the columns of Dataset 1*



PP Score between the columns of Dataset 1

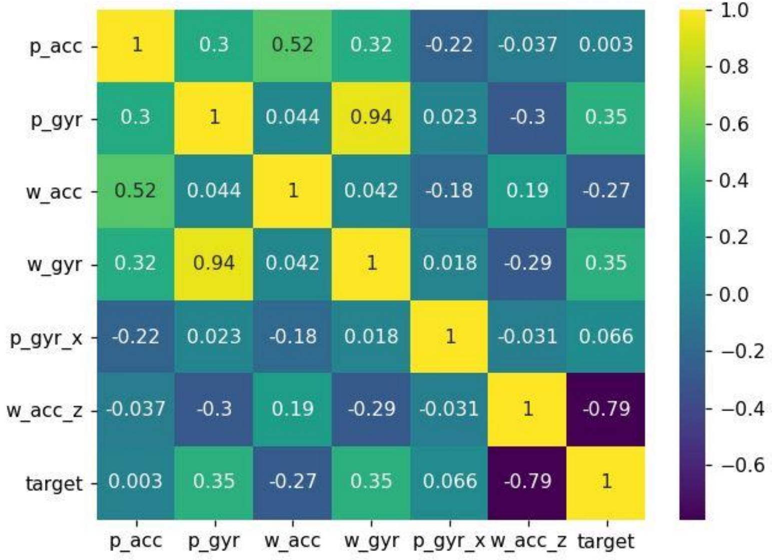


Dependencies heatmaps

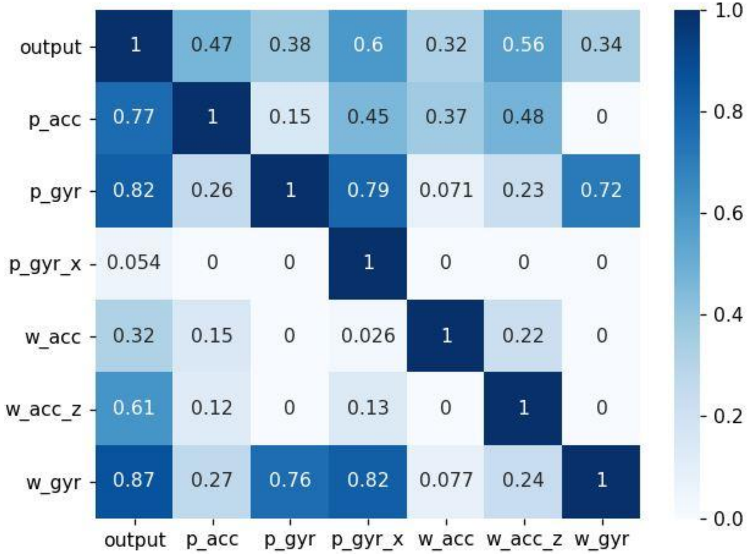
STISEN:

SECOND LEVEL FUSION

Correlation between the columns after first level of fusion of Dataset 1



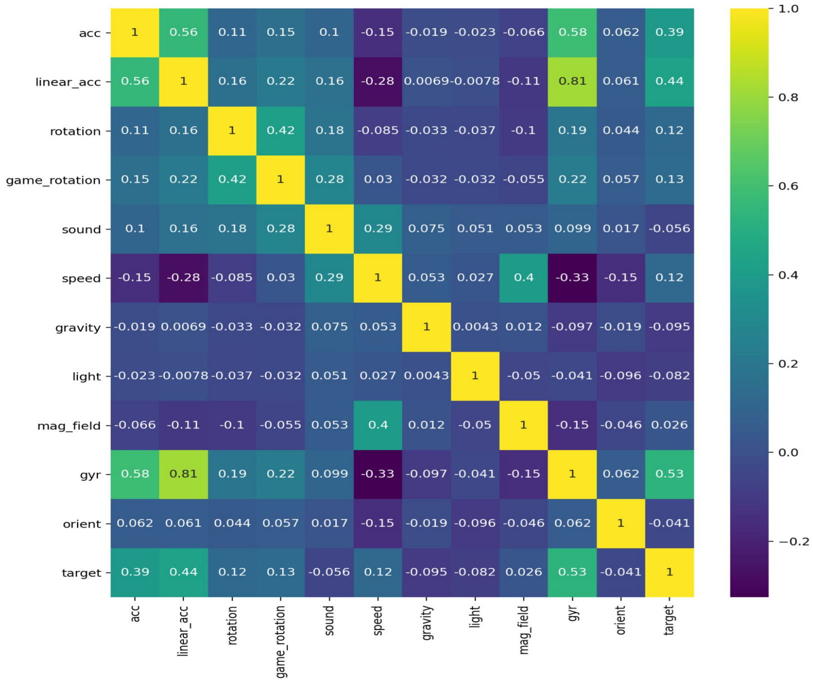
PP Score between the columns after second level of fusion of Dataset 1



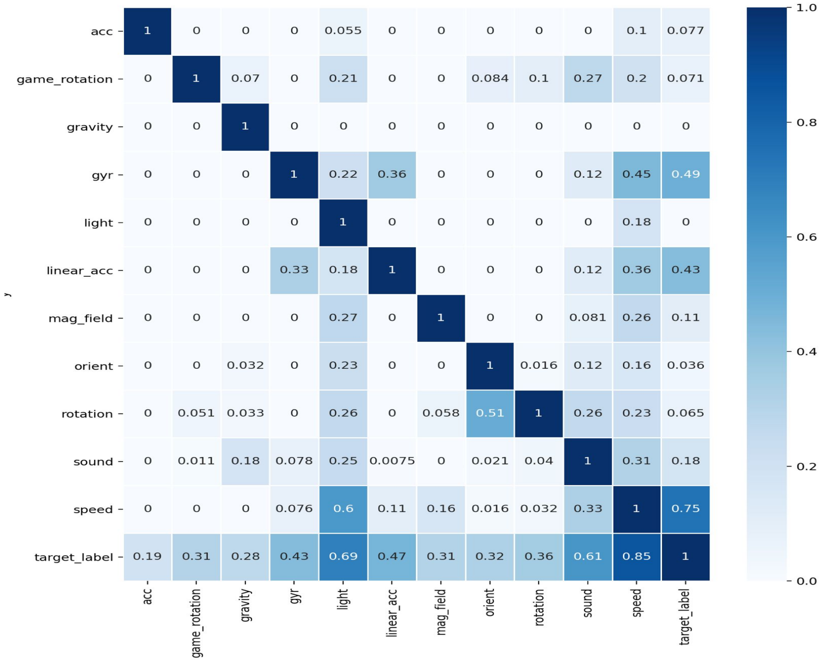
Dependencies heatmaps

TMD:

Correlation between the columns of Dataset 2

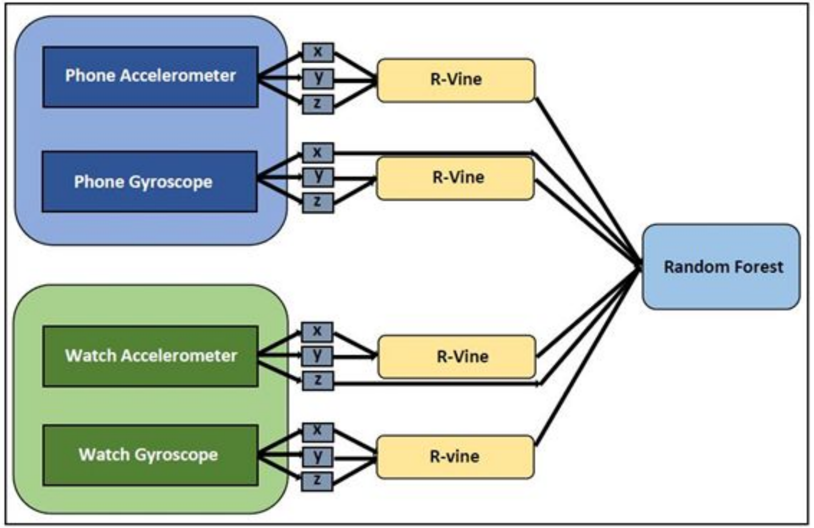


PPScore between the columns of Dataset 2

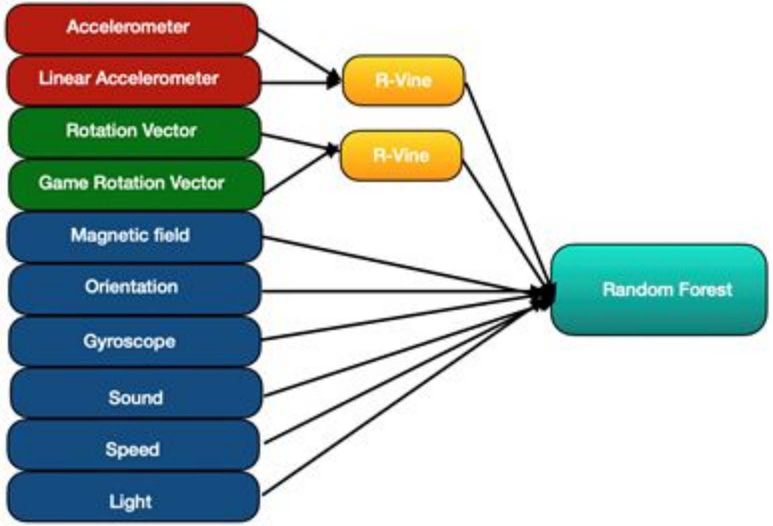


Architectures

STISEN:



TMD:



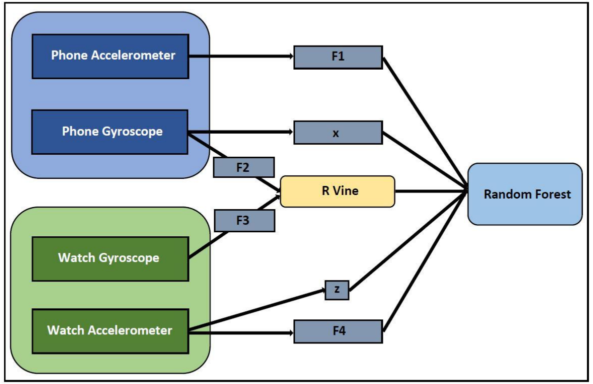
Fusion Algorithm

- The features that are required to be fused are given to R-Vine copula to determine the tree structure.
- The tree structure is generated based on Prim's maximum spanning tree algorithm.
- The marginal distributions of each of the features is estimated using KDE.
- The uniform marginal distributions of each of the features is estimated by applying probability integral transform.
- A bivariate copula is fit to each pair of features at each level of the tree based on ****** .
- The probability density of each fitted copula is computed.
- The joint density is obtained by using the pair copula construction equation.

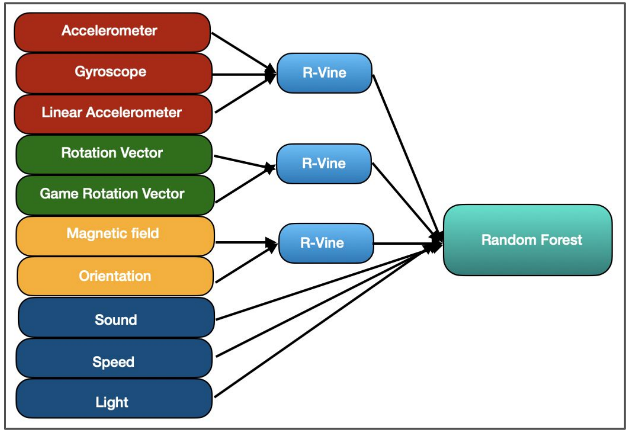
$$
\begin{aligned}
f(x_1,x_2,x_3) = {} & f_1(x_1) \times f_2(x_2) \times f_3(x_3) && (\textit{marginals}) \\
& \times c_{12}(F_1(x_1), F_2(x_2)) \times c_{23}(F_2(x_2), F_3(x_3)) && (\textit{unconditional pairs}) \\
& \times c_{13|2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3|x_2)) && (\textit{conditional pairs})
\end{aligned}
$$

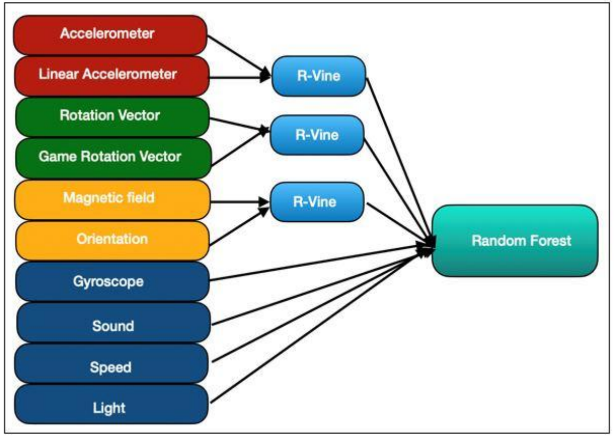
Other Architectures

STISEN:



TMD:





Results

STISEN:

Model	F1
First level fusion	98.50
Second level fusion of phone gyroscope and watch gyroscope	97.56

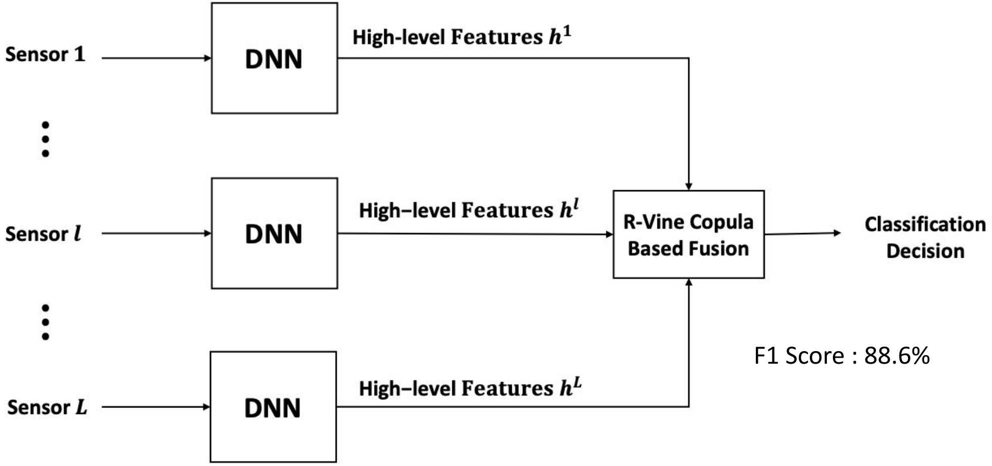
TMD:

Model	F1
Fuse [Acc, Linear Acc and Gyr] [Rotation and Game rotation][Magnetic field and orientation]	98.484
Fuse [Acc, Linear Acc] [Rotation and Game rotation][Magnetic field and orientation]	98.569
Fuse [Acc, Linear Acc] [Rotation and Game rotation]	98.643

Discussion

STISEN:

[2] S. Zhang, B. Geng, P. K. Varshney and M. Rangaswamy, "Fusion of Deep Neural Networks for Activity Recognition: A Regular Vine Copula Based Approach," 2019 22th International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, 2019, pp. 1-7

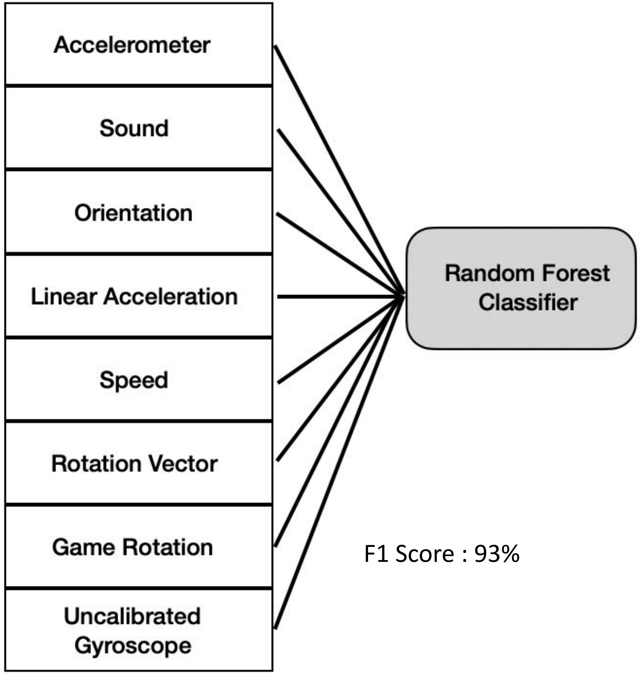


F1 Score : 88.6%

Discussion

TMD:

[5] C. Carpineti, V. Lomonaco, L. Bedogni, M. D. Felice and L. Bononi, "Custom Dual Transportation Mode Detection By Smartphone Devices Exploiting Sensor Diversity," 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2018, pp. 367-372, doi: 10.1109/PERCOMW.2018.8480119.



F1 Score : 93%

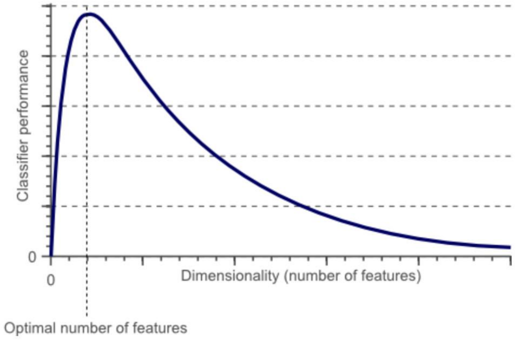
Discussion

Observations:

1. Various researches have shown that copula-based fusion of multiple sensing observations can significantly improve the performance of interference problems which is again seen in this work.

2. Feature selection plays a vital role in increasing the performance of the model.

3. Fusing Non-Dependent features for reducing dimension can decrease the accuracy. But the rate of decrease in accuracy is low.



* *medium.com*

Discussion

Findings:

Various structures have been experimented with to find the threshold value(+-0.4 : corr , 0.5 : PPS) for fusion of features. For the dependencies with the target (+-0.52, 0.56:PPS).

Future Work:

We wish that this (Finding the threshold for a given dataset) can be automated in future using statistical techniques. We also wish to find a way to get the optimal dimension for a given dataset.

Project timeline

Reviewing Fusion and representation Methods

Simulation of fusion techniques using Copulas and DNN on synthetic dataset.

Modeling different architecture s

12/09

05/10

4/11

1/12

Next

Literature review of previous applications of Multi-Modal Data

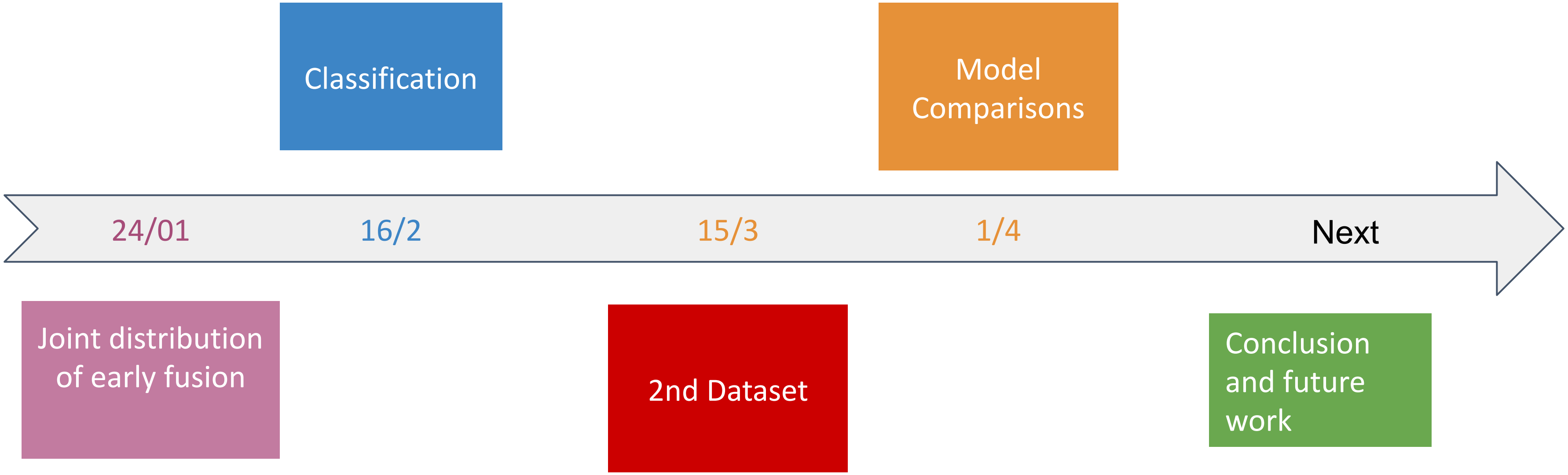
Collecting and Preprocessing Datasets

Review Paper

Modeling representation technique

Evaluation and comparison

Project timeline



Thank You…